



Demonstrating microbial co-occurrence pattern analyses within and between ecosystems

Ryan J. Williams¹, Adina Howe^{2,3} and Kirsten S. Hofmockel^{1*}

¹ Department of Ecology, Evolution, and Organismal Biology, Iowa State University, Ames, IA, USA

² Mathematics and Computer Science, Argonne National Laboratory, Argonne, IL, USA

³ Microbiology and Microbial Genetics, Michigan State University, East Lansing, MI, USA

Edited by:

Jay T. Lennon, Indiana University, USA

Reviewed by:

Albert Barberan, University of Colorado, USA

Joshua A. Steele, Southern California Coastal Water Research Project, USA

Cheryl-Emiliane Chow, University of British Columbia, Canada

*Correspondence:

Kirsten S. Hofmockel, Department of Ecology, Evolution, and Organismal Biology, Iowa State University, 251 Bessey Hall, Ames, IA 50011, USA

e-mail: khof@iastate.edu

Co-occurrence patterns are used in ecology to explore interactions between organisms and environmental effects on coexistence within biological communities. Analysis of co-occurrence patterns among microbial communities has ranged from simple pairwise comparisons between all community members to direct hypothesis testing between focal species. However, co-occurrence patterns are rarely studied across multiple ecosystems or multiple scales of biological organization within the same study. Here we outline an approach to produce co-occurrence analyses that are focused at three different scales: co-occurrence patterns between ecosystems at the community scale, modules of co-occurring microorganisms within communities, and co-occurring pairs within modules that are nested within microbial communities. To demonstrate our co-occurrence analysis approach, we gathered publicly available 16S rRNA amplicon datasets to compare and contrast microbial co-occurrence at different taxonomic levels across different ecosystems. We found differences in community composition and co-occurrence that reflect environmental filtering at the community scale and consistent pairwise occurrences that may be used to infer ecological traits about poorly understood microbial taxa. However, we also found that conclusions derived from applying network statistics to microbial relationships can vary depending on the taxonomic level chosen and criteria used to build co-occurrence networks. We present our statistical analysis and code for public use in analysis of co-occurrence patterns across microbial communities.

Keywords: co-occurrence, microbial communities, network theory, community assembly, MGRAST

INTRODUCTION

Co-occurrence relationships are ecologically important patterns that reflect niche processes that drive coexistence and diversity maintenance within biological communities (Tilman, 1982; HilleRisLambers et al., 2012). In microbial systems, niche processes like environmental filtering where abiotic factors define specific habitat limits can support coexistence (Horner-Devine et al., 2007; Costello et al., 2009; Ofiteru et al., 2010; Langenheder and Székely, 2011; Stegen et al., 2012), which are illustrated by co-occurrence patterns within communities. Species pairs or assemblages that co-occur may share similar ecological characteristics (Leibold and McPeck, 2006; Fuhrman and Steele, 2008; Raes and Bork, 2008; Chaffron et al., 2010; Eiler et al., 2012), which can be used to infer life-history strategies (Freilich et al., 2010; Barberán et al., 2012) and possibly to identify traits or even culture poorly understood microorganisms (Duran-Pinedo et al., 2011; Faust and Raes, 2012; Sun et al., 2013). Thus, applying co-occurrence analyses to microbial systems can provide valuable information for characterizing the biogeography, functional distribution or ecological interactions of microbes at the community scale or for identifying ecological traits of taxa that co-occur with well-characterized microorganisms.

Analyses of microbial co-occurrence patterns have been applied to a variety of research questions regarding biological interactions between organisms. Co-occurrence relationships have been useful in elucidating coexistence patterns spanning from pairs of microbial taxa in a range of ecosystems (Eiler et al., 2012; Kittelmann et al., 2013; Zhelnina et al., 2013) and functional groups (Duran-Pinedo et al., 2011; Bowen et al., 2013) to plant-microbe interactions (King et al., 2012). Classically, co-occurrence analysis has used checkerboard scores based on the presence or absence of organisms (Stone and Roberts, 1990), while larger datasets have been explored using correlation coefficients to represent either coexistence or competitive exclusion between two microbial taxa (e.g., Kittelmann et al., 2013). Subsequently, co-occurring pairs of microorganisms have been visualized using network methods (e.g., Fuhrman and Steele, 2008; Barberán et al., 2012) or ordination techniques [non-metric multidimensional scaling (NMDS)] as seen in King et al. (2012). Though these visualization methods are useful, there are very few examples of applying network statistics to microbial co-occurrence despite their growing popularity among subfields of ecological and evolutionary research (Proulx et al., 2005). Network statistics can be used to determine the importance of microorganisms in co-occurrence networks (e.g., degree,

betweenness, measures of centrality), possibly identifying key-stone species within an ecosystem (Bauer et al., 2010; Steele et al., 2011; Eiler et al., 2012). Additionally, little effort has been made to identify a multivariate test for differences in microbial community co-occurrence patterns between ecosystems. Coupling co-occurrence patterns within microbial communities to network or multivariate methods can enhance interpretation and therefore increase knowledge related to microbial co-occurrence.

The integration of a variety of analyses that have been used to study microbial co-occurrence patterns can allow researchers to understand microbial coexistence at multiple levels of biological organization. For example, the use of bivariate regressions, network statistics, and multivariate tests can be used to understand microbial co-occurrence between microbial pairs, within groups of co-occurring microorganisms (e.g., modules), and whole communities, respectively. We developed an approach that integrates these methods and then used multiple datasets to demonstrate our approach. While many of these approaches have been used previously, our analytical framework integrates several methods and applies multivariate statistics to test for differences in co-occurrence across ecosystems. We have also tested the robustness of our framework by including multiple taxonomic levels and considering alternative criteria for the construction of co-occurrence networks. Our analysis was implemented to answer the following co-occurrence-related research questions: (1) Are co-occurrence patterns among microbial communities the same among ecosystems? (2) Within communities, are there distinct modules of co-occurring microorganisms, and are these consistent among ecosystems? (3) Are pairs of co-occurring microbes consistent among ecosystems, and can ecological traits be inferred from these relationships? (4) Do these co-occurrence relationships change at different taxonomic levels or with various criteria used to construct co-occurrence networks? To test this approach, we used three publicly available datasets from the Metagenomics Analysis Server (MGRAS; Meyer et al., 2008). We expected to find that the majority of co-occurrence relationships would differ strongly across ecosystems creating vastly different modules of interacting taxa within each ecosystem, while potentially a few relationships will exist between pairs of microorganisms as a reflection of biological interactions that are present independently of environmental factors.

MATERIALS AND METHODS

We designed a statistical approach written in R v. 3.0.1 (R Core Team, 2013). All scripts necessary to replicate this analysis are included in the Supplementary Material. The analysis presented in this paper is designed to test for differences in co-occurrence patterns at the community level across ecosystems, identify modules of co-occurring microorganisms within communities, and identify pairwise co-occurrence patterns within modules that are consistent across ecosystems (summarized in **Figure 1**). We considered co-occurrence to be positive rank correlations (Spearman's correlation) between pairs of microbes within each dataset with the strength of the relationship represented by the correlation coefficient (**Figure 1B**). Negative correlations (indicative of either competitive interactions or non-overlapping niches between microbes; Faust and Raes, 2012) were also included in

this analysis though they were a small subset of our combined datasets. We only considered negative and positive co-occurrence relationships based on strength of correlation (i.e., ρ from the Spearman's correlation) at values less than or equal to -0.75 and -0.5 or greater than 0.5 and 0.75 .

We applied our approach to determine co-occurrence patterns from three public datasets maintained through MGRAS that had replicated samples of 16S rRNA amplicon sequencing. Abundances of classified bacteria and archaea were accessed using the *matR* package (Braithwaite and Keegan, 2013), and were summarized at the order and family level with the assumption that microorganisms share similar traits at these phylogenetic levels. Though there is some evidence that certain traits are conserved at high levels of phylogeny (Philippot et al., 2010), we tested our analysis at multiple taxonomic levels as coherence of ecological patterns like co-occurrence may vary across different levels of taxonomy (Koeppel and Wu, 2012). The datasets were grouped following the schematic in **Figure 1A** with replicates nested within ecosystems. Ecosystems included apple flowers with and without antibiotic application [Shade et al., 2013b; 2 flower types (replicates) with 15 samples of each type], human body surfaces [Costello et al., 2009; 9 different bodies (replicates) divided into males and females with 24–25 samples each], and soils from different land-use types [Lauber et al., 2008; 5 different soils (replicates) with 4–43 samples each]. Datasets were chosen based on the number of replicates nested within similarly sampled ecosystems (i.e., flowers, body surfaces, or soils), and were classified generally into different replicates within each ecosystem. While the classification of these samples may not represent ideal replicates from each study [e.g., communities differ across body surfaces rather than sex or individual in the study by Costello et al. (2009) and communities did not differ across flower antibiotic treatments (Shade et al., 2013a)], they do provide enough statistical power to demonstrate our approach. Thus, it should be noted that biological interpretation of our results requires further exploration through controlled studies.

Before beginning our analysis, we rarefied samples to standardize for sequencing depth between samples. Prior to rarefaction, samples ranged between 2 and 12,000 sequences per sample and a mean ranging from 1000 to 5000 depending on ecosystem type; these values were similar across taxonomic levels. We chose to use the minimum amount of counts per sample from the Shade et al. (2013b) datasets as this number was roughly the average for all samples used in our analysis. However, this rarefaction step led us to using only two different soils from Lauber et al. dataset (2008) and three female body datasets from Costello et al. (2009). Though this rarefaction step reduced the number of datasets used, it also removed less abundant taxa that can produce spurious co-occurrence relationships with highly abundant taxa (Faust and Raes, 2012). For the order dataset, samples were rarefied to 1407 reads per sample while the family dataset was rarefied to 1353 reads per sample.

TESTING FOR DIFFERENCES IN CO-OCCURRENCE PATTERNS AT THE COMMUNITY LEVEL

To test for differences in co-occurrence patterns between microbial communities from different ecosystems, we generated a

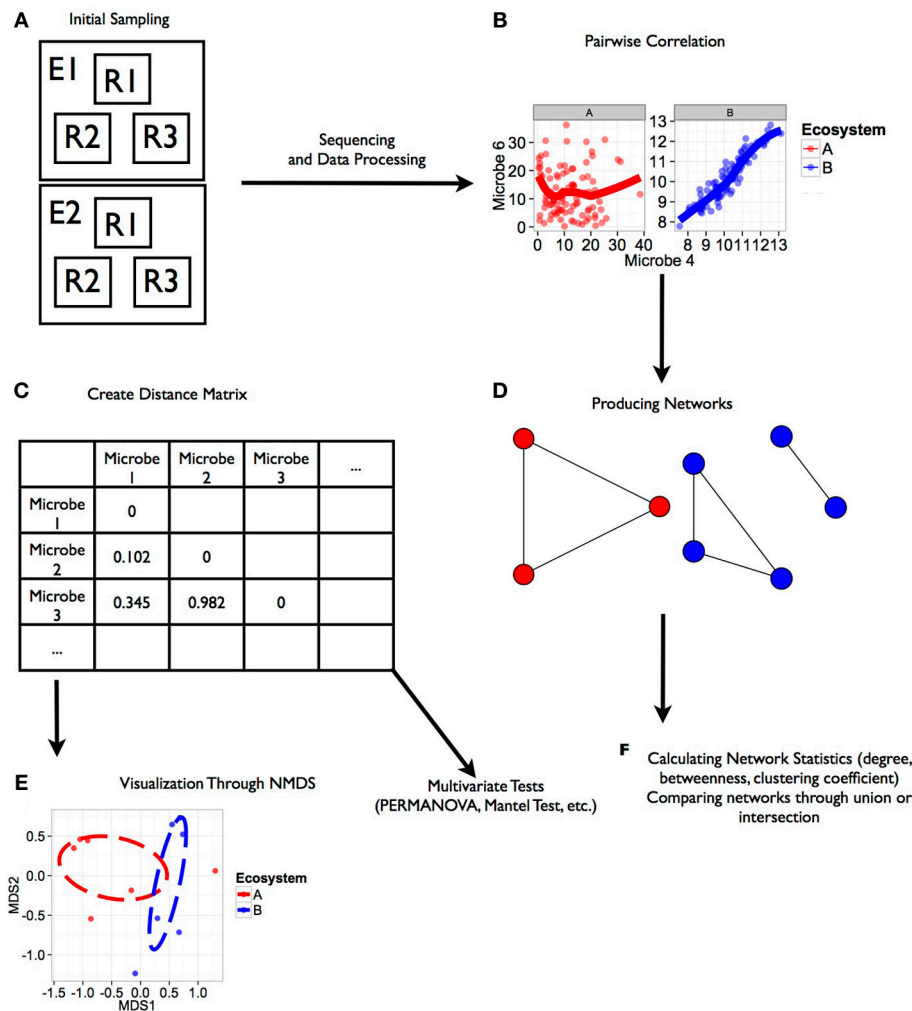


FIGURE 1 | Workflow for analysis of microbial co-occurrence between ecosystems. This illustration represents a workflow from data collection through analysis stages for determining co-occurrence patterns among microbial communities. Each step in the workflow has been generated from simulated data. Scripts for the generating these figures are located in the Supplemental Material. **(A)** Ecosystems were sampled (E_1 , E_2), and within each ecosystem several replicate groups of random samples were taken (R_1 , R_2 , R_3). **(B)** Rank correlation represented by this regression plot was performed for two microbial orders (Microbe 4 and 6 shown here) within each environment that were consistent among replicate groups. **(C)** Distance

matrices based on correlation coefficients between taxa were generated for downstream statistical tests. **(D)** Ecosystem-specific co-occurrence patterns were visualized using network diagrams. **(E)** Co-occurrence relationships between each ecosystem were visualized using NMDS. Further tests of network topology and distance matrices can be performed using a variety of multivariate tests like the mantel test or permutation multivariate analysis of variance (PERMANOVA). In the case of our simulated data, we found a significant effect of ecosystem on co-occurrence (PERMANOVA; $P < 0.02$). **(F)** Additional network statistics can be calculated to characterize networks, and networks can be compared to find shared relationships.

dissimilarity matrix consisting of Spearman correlation coefficient distances (1-correlation coefficient) representing co-occurrence between all pairs of microorganisms from each sample (**Figure 1C**) using the bioDist package (Ding et al., 2014). The calculation of these distances produces a matrix where microbial taxa rather than samples were compared to one another. This Spearman's distance matrix represents the strength of correlation among microbial pairs; thus smaller distances represent stronger correlations, which were visualized using non-metric multidimensional scaling (NMDS; **Figure 1E**). We used a permutational multivariate analysis of variance (PERMANOVA; 9999 permutations) (Anderson, 2001) from the vegan package (Oksanen

et al., 2013), with ecosystem type (apple flower, bodies, or soils) representing our independent variable to test for differences in co-occurrence patterns at the community level based on the Spearman's distance matrix.

The generation of this Spearman's dissimilarity matrix and its use in a PERMANOVA has not been described previously to our knowledge; therefore we generated simulations under a variety of conditions that represent null cases and significant differences in community co-occurrence patterns between ecosystems (R script in Supplementary Material). The null case represents a situation where correlations between two microorganisms within a community are no greater than any correlation

with a microorganism sampled from another ecosystem, where no correlation is expected. If correlations between microorganisms within a community were strong and consistent across replicates from the same ecosystem, this null hypothesis would be rejected (Supplementary Figure 1).

DELINEATING MODULES OF CO-OCCURRING MICROORGANISMS AND CONSISTENT CO-OCCURRENCE RELATIONSHIPS

We illustrated modules of co-occurring microorganisms within communities where microbial taxa represent nodes and the presence of a co-occurrence relationship based on correlation is represented by an edge (Figure 1D). These correlation relationships were generated for each pair of microbial taxa within each ecosystem replicate as long as both taxa had abundance greater than 0. We made a consensus network of co-occurrence relationships within each ecosystem based on the strength of the correlation (ρ from the Spearman's correlation), and co-occurrence relationships were only included if they occurred across all ecosystem replicates. Though this method has been illustrated to produce some spurious co-occurrence relationships among simulated data (Friedman and Alm, 2012), this rank-based correlation statistic does not require any transformation of variables to fit assumptions of normality and may outperform Pearson's correlations. To increase our level of stringency that may reduce the appearance of spurious co-occurrences within our networks, pairwise relationships had to be consistent across all datasets of a given ecosystem type, greatly reducing the number of co-occurrence pairs (Chaffron et al., 2010).

Networks were produced using the *igraph* package (Csardi and Nepusz, 2006) where each network was the union of positive co-occurrences or negative co-occurrences (less than -0.5 or greater than 0.5) that were consistent within each ecosystem. Unconnected nodes were removed along with loops that indicate microbial taxa were correlated with themselves using the "delete.vertices" and "simplify" functions, respectively. We performed this through the "graph.union.by.name" function from the *igraph* package. Modules were designated as groups of highly connected microbes (modules) that were poorly connected to others. Modules were detected using an algorithm based on edge betweenness through the "edge.betweenness.community" function in *igraph* (Girvan and Newman, 2002; Newman and Girvan, 2004). The method used in our analysis looks for edges (i.e., co-occurrence) that are the most between vertices (microbes), and thus finding edges that are responsible for connecting many other microbial groups (Girvan and Newman, 2002). This method differs from agglomerative methods [e.g., measures of "cliquishness" (Watts and Strogatz, 1998)], which have been demonstrated in protein-network clustering (Bader and Hogue, 2003; Rivera et al., 2010). Instead, the betweenness centrality method we use is designed for simple graphs with single-type vertices as opposed to bipartite graphs, and avoids hierarchical clustering issues that can occur with agglomerative methods (Girvan and Newman, 2002). We also looked for intersections between networks from different ecosystems using the "graph.intersection.by.name" function (*igraph*) to determine if any co-occurrence relationships were consistent across ecosystems.

ADDITIONAL STATISTICAL ANALYSES

To characterize differences in community composition between ecosystems, we performed a PERMANOVA with Bray–Curtis dissimilarity on our initial community matrices (for both microbial orders and families) with abundances scaled between 0 and 1. This analysis was performed using the "decostand" and "adonis" functions from the *vegan* package in R (Oksanen et al., 2013). We generated nonmetric multidimensional scaling (NMDS) plots to visualize differences in community composition using Bray–Curtis dissimilarity as well.

We were also interested in generating statistics that describe the network that may be important for understanding co-occurrence relationships. We produced network statistics that describe the position and connectedness of microorganisms within each co-occurrence network. This included normalized node degree, which is the number of co-occurrence relationships that a microorganism is involved in a network normalized by the total number of nodes using the "degree" function (*igraph* package; Csardi and Nepusz, 2006). We also calculated betweenness scores for each microbial taxonomic group using the "betweenness" function from *igraph* (Csardi and Nepusz, 2006), which is defined by the number of paths through a focal microbial node. Additionally, we calculated clustering coefficients using the "transitivity" function for comparison to other networks as performed in Steele et al. (2011).

We then determined relationships between degree and betweenness. Initial visualization of relationships betweenness and degree appeared to be correlated and non-linear. Thus we fit mixed models within each ecosystem and each level of correlation strength with degree as an independent variable, betweenness as a response variable, and ecosystem replicate as a random factor based on a power function (αx^β). Mixed models were fit using the *lme4* package in R (Bates et al., 2014). With this analysis we hoped to identify microbial taxa that are highly connected that may represent keystones within their ecosystem (Steele et al., 2011; Faust and Raes, 2012). We expanded this concept of keystone species to include both degree and betweenness, as these metrics illustrate both the number of connections and how important those connections are to the overall network. Therefore, we identified keystone taxa as those with the highest predicted betweenness based on our mixed models.

RESULTS

DIFFERENCES IN CO-OCCURRENCE PATTERNS AT THE COMMUNITY LEVEL

We first quantified differences in community composition and community co-occurrence across ecosystems using a PERMANOVA and the Bray–Curtis dissimilarity and Spearman's distance, respectively. Although differences in community composition were clear among microbial orders and families (Supplemental Figure 1, $P < 0.0001$ for both), no clear difference was seen in co-occurrence patterns ($P > 0.05$). The lack of differences was clear in the visualization through NMDS as samples from each ecosystem completely overlapped one another (data not shown). The lack of differences in community co-occurrence patterns were likely driven by weak or non-significant correlations between most taxa within each ecosystem (see Supplementary

Material for simulation of this case). Thus, our approach did not detect differences between co-occurrence patterns between samples from different ecosystems. In other words, the majority of microorganisms within a single ecosystem replicate were uncorrelated, and therefore equally uncorrelated to microorganisms from any other ecosystem replicate. If stronger correlations existed within a single ecosystem replicate as compared to other unrelated replicates, the explanatory power of this analysis would increase (see Supplementary Material).

DELINEATING CO-OCCURRING MODULES AND PAIRS

After testing for differences in community co-occurrence patterns between ecosystems, we aimed to identify consistent groups or modules of co-occurring microbial taxa among replicate samples within an ecosystem (Figure 2; Supplementary Tables 2, 3). When considering microbial orders, the apple ecosystem had the most modules at 11 followed by male samples with 4 and female and soil both with 3. When classifying microbial families into modules, a different trend was found. Soil had the most modules at 18, followed by apple at 14, female with 7, and male with 5. Negative co-occurrence modules were not found in any of the

body samples (male or female), while soil had the most (9 order modules, 7 family modules) and apple had only a few (3 order, 4 family). In general, modules contained between either 2–6 orders or families, and each ecosystem usually had one large module containing multiple taxa. For example among soil families, one module contained 41 taxa while other soil family modules contained between 2 and 10 taxa. Modules were often found to be composed of multiple unrelated bacterial orders or families that were not necessarily associated at higher taxonomic levels. Thus, module delineation did not necessarily follow phylogenetic relationships among microbial communities categorized at the level of orders or families.

We then aimed to determine pairwise co-occurrence relationships that were consistent across ecosystems through the intersection of networks from different ecosystems (Table 1). Overall, more microbial families co-occurred across ecosystems than microbial orders, and no co-occurrence relationships held across all ecosystems. Also, relationships found at one taxonomic level were not necessarily found at another level. For example, Cytophagales and Flavobacteriales co-occurred across soil and apple ecosystems, and this relationship

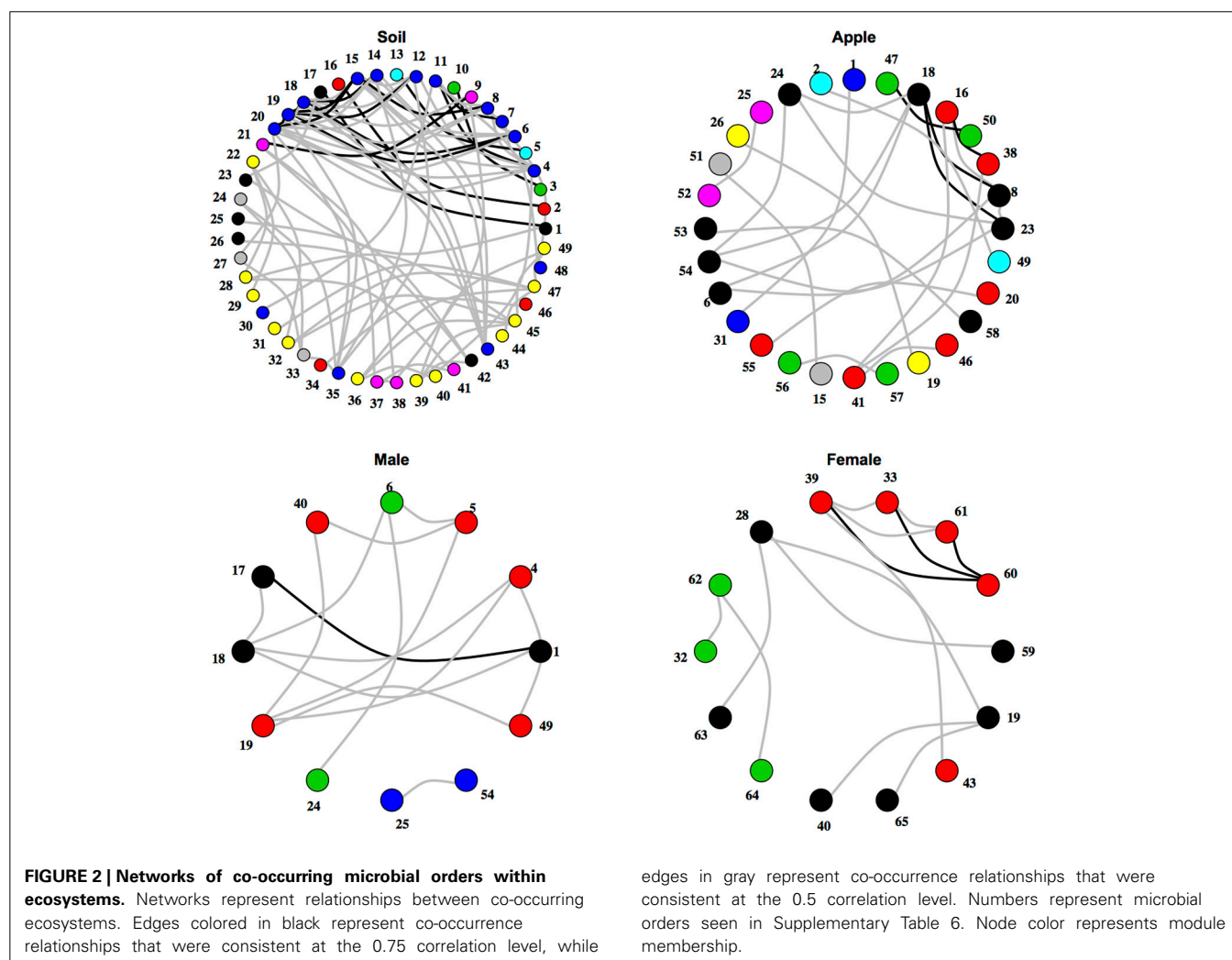


Table 1 | Pairwise co-occurrence relationship statistics.

Ecosystem comparison	Orders	Families
Soil—Apple	Cytophagales—Flavobacteriales Cytophagales—Sphingobacteriales	Clostridiaceae—Mycobacteriaceae Cytophagaceae—Flavobacteriaceae Cytophagaceae—Oxalobacteraceae Cytophagaceae—Propionibacteriaceae Cytophagaceae—Sphingobacteriaceae Microbacteriaceae—Micrococcaceae Microbacteriaceae—Propionibacteriaceae Microbacteriaceae—Pseudonocardiaceae Microbacteriaceae—Sphingobacteriaceae Micrococcaceae—Nitrosomonadaceae Micrococcaceae—Propionibacteriaceae Micromonosporaceae—Promicromonosporaceae Propionibacteriaceae—Pseudonocardiaceae Propionibacteriaceae—Sphingobacteriaceae Rhodocyclaceae—Rhodothermaceae
Soil—Male	Acidimicrobiales—Solirubrobacterales Burkholderiales—Sphingobacteriales Cytophagales—Sphingobacteriales	Acidimicrobiaceae—Conexibacteraceae Cytophagaceae—Nocardoidaceae Microbacteriaceae—Oxalobacteraceae Oxalobacteraceae—Rhodobacteraceae
Soil—Female		Microbacteriaceae—Propionibacteriaceae Microbacteriaceae—Sphingomonadaceae
Apple—Male		Cytophagales—Sphingobacteriales Propionibacteriaceae—Sphingomonadaceae
Apple—Female		Clostridiales Fam. XI <i>Incertae Sedis</i> —Corynebacteriaceae Microbacteriaceae—Propionibacteriaceae
Male—Female	Pseudomonadales—Sphingomonadales	Corynebacteriaceae—Mycobacteriaceae Moraxellaceae—Pseudonocardiaceae Moraxellaceae—Sphingomonadaceae

Pairs of microbial taxa represent consistent positive pairwise relationships across the designated ecosystems.

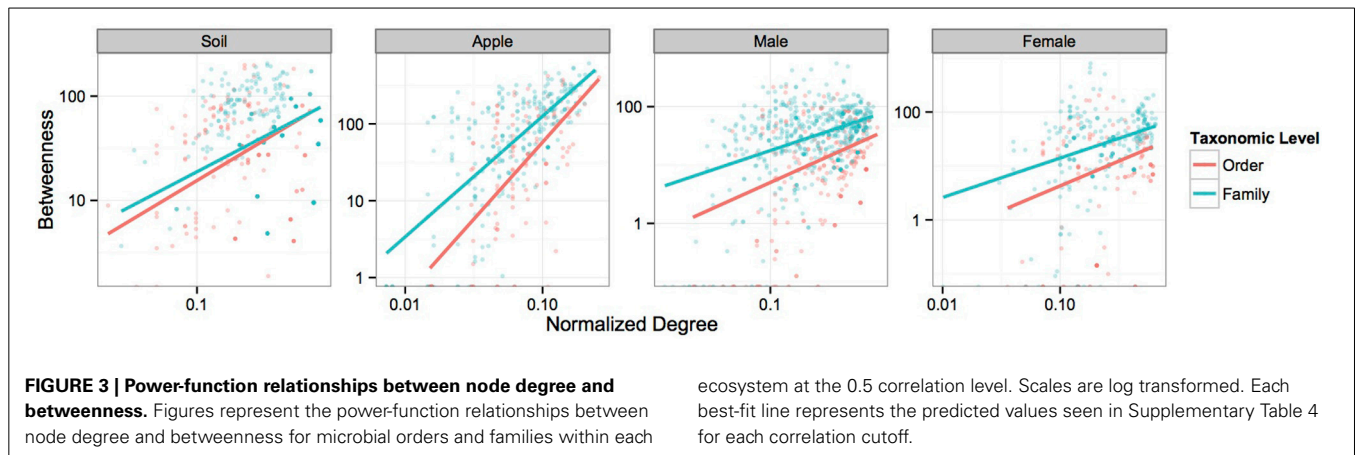
held true between Cytophagaceae and Flavobacteriaceae. Alternatively, Micrococcaceae from the Actinomycetales and Nitrosomonadaceae from the Nitrosomonadales co-occurred at the family level, but their respective orders did not co-occur. Furthermore, important co-occurrence relationships among families within the same order, such as Micrococcaceae and Microbacteriaceae from the Actinomycetales, were not detectable when considering microbial order alone.

CO-OCCURRENCE NETWORK STATISTICS

We first visualized networks within each ecosystem for both positive and negative co-occurrence relationships (Figure 2, Supplemental Figure 2). We then calculated a normalized degree and betweenness score for nodes within each network and modeled relationships between these variables as a power function, αx^β , using mixed models. The slopes of each power function within an ecosystem were similar across taxonomic levels when considering correlations greater than 0.05 (Figure 3). However, when considering more stringent correlation cutoffs, greater

disparity was seen across power functions within an ecosystem (Supplementary Figure 3), suggesting that the choice of taxonomic level or correlation strength may have a significant effect on the interpretation of co-occurrence networks. All but two cases had significant slope parameters (β ; Supplementary Table 4), and involved correlation cut offs of either 0.75 or -0.75 . When considering the slopes across different strengths of correlation, models based on negative co-occurrence networks often produced higher values of β ; this was especially true when considering correlations less than or equal to -0.5 . We also calculated clustering coefficients for comparison to other biological networks. We found that while all networks fell across a range of values common to other networks (Steele et al., 2011), only positive co-occurrence networks displayed “small-world” characteristics (Watts and Strogatz, 1998), where nodes were more connected on average than may be expected at random (Supplemental Table 5).

We then used the predictions of our mixed models to determine keystone taxa within each network of positive co-occurrences. When considering networks with different



correlation cutoffs within a similar ecosystem, the top keystone taxa were not necessarily the same. In soils, Bacillales, Actinomycetales, and Clostridiales were the top keystone orders with a cutoff of 0.75 while Thermoleophilales, Desulfotomaculales, and Sphingobacteriales were designated as the top keystone orders with a cutoff of 0.5. However when moving down to the family level in soils, keystone taxa were much more consistent between networks with different correlation cutoffs. These results suggest that applying ecological characteristics to network elements must happen under careful consideration of the parameters used to delineate co-occurrence relationships.

DISCUSSION

The exploration of co-occurrence networks is a useful method for determining biological interactions occurring within microbial communities. Here we have laid out a framework to generate co-occurrence networks and to compare co-occurrence relationships within and between ecosystems. A novel strength of this framework is its utility at multiple scales; analysis can be performed to observe co-occurrence from the community level down to pairwise interactions between microbial taxa. We applied our analyses to three datasets to demonstrate its effectiveness and determine differences in co-occurrence between ecosystems. Through this investigation, we were able to distinguish co-occurring pairs of microbial orders and families that were consistent across ostensibly different ecosystems, while the majority of co-occurrence relationships within ecosystems appear to be at random (i.e., uncorrelated microbial pairs). Additionally, we were able to distinguish modules of co-occurring microorganisms that appear to behave similarly within communities. These results and our approach can be used to explore microbial communities in a variety of ecological contexts including but not limited to the identification of biotic and abiotic drivers of microbial community assembly, identification of keystone microbial species, or inferring ecological characteristics of poorly understood or unculturable microbial taxa.

The analytical framework that we present has been able to detect ecologically relevant relationships between microbial taxa. For example, we were able to detect consistent positive co-occurrence between two skin-dwelling bacteria, Pseudomonadales and Sphingomonadales, across male and

female body datasets. One important use of our analytical framework is the development of hypotheses regarding traits of rarely studied microbes through co-occurrence with other microorganisms based on the assumption that coexisting species are ecologically similar (Leibold and McPeck, 2006; Barberán et al., 2012). For example, the recently described order, Solirubrobacterales, has been noted to occur in soils with little information regarding its ecological role (Shange et al., 2012). Our co-occurrence analysis suggests that Solirubrobacterales either assumes analogous ecological roles or is selected by similar environmental factors as its co-occurring taxa in soil (Figure 2). Strains from Acidimicrobiales and Actinomycetales are known to overlap in their carbon substrate use (Goldfarb et al., 2011). These results illustrate potential resource utilization roles that minimize interspecific competition through niche partitioning, where Solirubrobacterales can coexist with Acidimicrobiales and Actinomycetales by utilizing alternative substrates. Alternatively, these three heterotrophic orders may have overlapping carbon substrate preference, yet competition between the three orders is minimized under C-rich soil conditions. Indeed, the relationship between Solirubrobacterales and Acidimicrobiales and the related families, Acidimicrobiaceae and Conexibacteraceae, may be ecologically relevant as these co-occurrence relationships occurred in both soil and male body ecosystems (Table 1). The relationships between these groups of microorganisms represent testable hypotheses regarding coexistence between newly described bacteria like the Solirubrobacterales and other microbial heterotrophs. Furthermore, hypotheses can address higher levels of hierarchical organization among co-occurring pairs by exploring relationships between microbial taxa with similar life history (e.g., heterotrophy) that exist within the same module indicating similar niches (Chow et al., 2013). All together, these relationships represent potential hypotheses driven by analysis through our co-occurrence approach and require the inclusion of more replicated microbial community data to confirm coexistence between these microbial taxa.

In microbial systems, much attention has been paid to the deterministic or stochastic assembly of communities. While stochastic processes may play a partial role in microbial community assembly, environmental filtering or selection by abiotic factors can be important in both experimental (Ofiteru

et al., 2010; Langenheder and Székely, 2011; Faust and Raes, 2012) and naturally occurring communities (Horner-Devine et al., 2007; Costello et al., 2009; Stegen et al., 2012). We used our analysis framework to test for differences in co-occurrence networks at the community level, and found that though community composition strongly differs between ecosystems (**Figure 2**), no significant differences existed among community co-occurrence. Rather, few co-occurrence relationships are strong within ecosystems, yet some of the co-occurrences are consistent across ostensibly different ecosystems. These results suggest that environmental filtering plays a strong role in driving microbial community composition and fluctuations among microbial populations are generally independent of one another. However, further examination of uncorrelated microbial populations across more ecosystems is necessary, as these datasets were not collected to explicitly test microbial co-occurrence and the scale at which samples were collected may not be relevant for microbial community interactions. It has been suggested that some microbial taxa may be more affected by biotic factors, while others are more affected by abiotic factors (Fuhrman and Steele, 2008), which may create complex patterns within co-occurrence networks that we could not detect with this method. Though our analysis was able to illustrate differences in co-occurrence at the community level among simulated data (Supplemental Material), the true data used in our analysis was much more complex and had less replications. Further application of PERMANOVA for co-occurrence may need to consider the amount of replication necessary to pick up differences in community co-occurrence among “noisy” natural data. Also, the incorporation of continuous environmental covariates may explain variation in co-occurrence or determine at least the abiotic effects on community co-occurrence (Steele et al., 2011) as it has already been used to forecast microbial community composition (Larsen et al., 2012). Nevertheless, these results indicate that the majority of biological interactions between microbial taxa are ecosystem dependent much like microbial temporal dynamics (Shade et al., 2013a), and consistent biological interactions among microorganisms may be a special case rather than the norm when considering microbial communities as a whole and at high taxonomic levels.

Though we were able to demonstrate the usability of our analytical framework and find potentially useful interactions between microbial taxa, there are a few shortcomings to what we present here. One aspect of our analysis that we did not test is the relative contribution specific ecosystem replicates may have on overall co-occurrence relationships. Unequal sample sizes among replicates is an experimental factor worth considering as the use of PERMANOVA and other multivariate tests can be sensitive to unbalanced designs (Anderson and Walsh, 2013). Also, the number of ecosystem replicates might affect our ability to detect consistent co-occurrence patterns. Our apple and soil datasets had two ecosystem replicates, while the body dataset consisted of three and six replicates for female and male bodies, respectively. With greater replication across all ecosystems, one might relax their criteria for determining consistent co-occurrence relationships and instead consider the distribution of correlation coefficients across replicates. Additionally, special consideration may be needed when choosing module-detection algorithms, and

comparisons between agglomerative (Watts and Strogatz, 1998; Rivera et al., 2010) and divisive methods as we used here (Girvan and Newman, 2002). Though it should be noted that the networks we analyzed were fairly simple and may not vary largely depending on the community detection method.

We also chose in our analysis to assemble networks based on correlation coefficients without consideration of the involved *p*-value. When considering correlation strength cutoffs, we produced different networks (**Figure 2**), and statistics like degree and betweenness calculated from these models were different as well. Therefore biological interpretation of these statistics may need to consider the sensitivity of these biological interpretations to changes in criteria determining network relationships. Similarly, we did test whether cut-offs based on *p*-values, or adjusted *p*-values based on false discovery rate (*q*-value; Strimmer, 2008) affected our results (data not shown). We observed that an adjustment based on false discovery rate actually produced *q*-values less than *p*-values based on pairwise correlations (Pike, 2011). It is important to note that each ecosystem and the datasets belonging to each ecosystem had varying sample size, which can also affect the *p*-value of the correlation. Despite differences in sample sizes among these data, and the variety of methods that exist today in analyzing networks, the results we have presented are a clear and accessible example of how our analytical framework for co-occurrence analysis allows for deep investigation of environmental factors and biological interactions occurring at multiple scales of biological organization. Co-occurrence relationships found in our study necessitate further observation across multiple datasets and empirical tests that determine the mechanisms driving co-occurrence between specific microorganisms.

The use of network algorithms and statistics to understand co-occurrence within communities can play an important role in understanding drivers of community assembly among microorganisms (Faust and Raes, 2012). Expanding previous research that focuses on bivariate comparisons of microbial taxa (e.g., Zhahnina et al., 2013) through the use of multivariate techniques as we have demonstrated here is an important next step. The statistical analyses that we provide can be applied to any sort of community abundance data, and is not necessarily limited to microbial applications. Additionally, alternative measures of co-occurrence like sparCC (Friedman and Alm, 2012), maximal information coefficient (MIC; Reshef et al., 2011) may be incorporated throughout the framework instead of Spearman's correlation. When moving to lower levels of taxonomic resolution like species, it may be important to incorporate measures like MIC which has been demonstrated to identify relationships with fine taxonomic resolution (Reshef et al., 2011). However, the actual biological interpretation at this scale may be difficult, even when utilizing methods like MIC due to the number of co-occurrence relationships and the paucity of ecological data regarding the majority of 16S rRNA sequences. Our analysis does not strictly require the use of Spearman's correlations, and other methods that measure the strength of a relationship between pairs of microbes can be easily incorporated. Additionally, the Spearman's distance may be changed by scaling any other measure (MIC, for example) between 0 and 1, subtracting that value from

1, and thereby creating a distance matrix that can be incorporated into a multivariate framework.

Despite some of the shortcomings presented here, the framework we present may also be useful in conjunction with other methods that measure phylogenetic dispersion while investigating community assembly (Walter and Ley, 2011), and are easily calculated using phylogenetic trees used or created in through sequencing pipelines (e.g., QIIME; Caporaso et al., 2010). Additionally, using genomic data that relate traits across wide spans of phylogeny (e.g., Zimmerman et al., 2013) or the combination of metagenomic data and phylogenetic relationships (e.g., Segata et al., 2012), may be used to validate ecological inferences based on co-occurrence. Linking these traits with modules of co-occurring microorganisms may be useful for identifying functional groups within communities, where modules rather than individual taxa may be used to simplify high-dimensional datasets. Furthermore, linking co-occurrence relationships with both traits and environmental metadata (Fuhrman, 2009; Steele et al., 2011; Gilbert et al., 2012) may be applied in our framework to test for effects of abiotic factors on multiple levels of co-occurrence. The calculations of additional network statistics can be performed at the node, edge, or network level like clustering coefficients (Steele et al., 2011), which can easily be incorporated into scripts included in the Supplementary Material. Though the applicability of our approach is broad, the results we present here a demonstration of our analytical framework and are also hypotheses meant for further investigation. Further application of co-occurrence analysis is necessary in reduced experimental systems to conclude that co-occurrence relationships found here are driven by biological or environmental factors (Gilbert et al., 2012), which in turn has proven successful in understanding uncultured microorganisms (Duran-Pinedo et al., 2011; Faust and Raes, 2012). Our co-occurrence framework represents a step toward understanding microbial ecology beyond community composition alone, and our analysis at multiple scales of biological organization can help us understand community assembly and coexistence among microorganisms (Raes and Bork, 2008; Fuhrman, 2009; Faust and Raes, 2012).

ACKNOWLEDGMENT

We would like to thank Elizabeth Bach, Brian Brunelle, Sarah Hargreaves, Alison King, Brian Wilsey, Fan Yang, and reviewers for helpful comments on versions of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00358/abstract>

REFERENCES

- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Aust. Ecol.* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x
- Anderson, M. J., and Walsh, D. C. I. (2013). PERMANOVA, ANOSIM, and the Mantel test in the face of heterogeneous dispersions: what null hypothesis are you testing? *Ecol. Monogr.* 83, 557–574. doi: 10.1890/12-2010.1
- Bader, G. D., and Hogue, C. W. V. (2003). An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* 4, 1–27. doi: 10.1186/1471-2105-4-2
- Barberán, A., Bates, S. T., Casamayor, E. O., and Fierer, N. (2012). Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J.* 6, 343–351. doi: 10.1038/ismej.2011.119
- Bates, D., Maechler, M., Bolker, B., and Walker, S. (2014). *lme4: Linear Mixed-effects Models Using Eigen and S4*. R package version 1.0-6. Available online at: <http://CRAN.R-project.org/package=lme4>
- Bauer, B., Jordán, F., and Podani, J. (2010). Node centrality indices in food webs: rank orders versus distributions. *Ecol. Complex* 7, 471–477. doi: 10.1016/j.ecocom.2009.11.006
- Bowen, J. L., Byrnes, J. E. K., Weisman, D., and Colaneri, C. (2013). Functional gene pyrosequencing and network analysis: an approach to examine the response of denitrifying bacteria to increased nitrogen supply in salt marsh sediments. *Front. Microbiol.* 4:342. doi: 10.3389/fmicb.2013.00342
- Braithwaite, D. T., and Keegan, K. P. (2013). *matR: Metagenomics Analysis Tools for R*. R package version 1.0.0.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335–336. doi: 10.1038/nmeth.f.303
- Chaffron, S., Rehrauer, H., Pernthaler, J., and Von Mering, C. (2010). A global network of coexisting microbes from environmental and whole-genome sequence data. *Genome Res.* 20, 947–959. doi: 10.1101/gr.104521.109
- Chow, C. E. T., Sachdeva, R., Cram, J. A., Steele, J. A., Needham, D. M., Patel, A., et al. (2013). Temporal variability and coherence of euphotic zone bacterial communities over a decade in the southern California bight. *ISME J.* 7, 2259–2273. doi: 10.1038/ismej.2013.122
- Core Team, R. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna. Available online at: <http://www.r-project.org/>
- Costello, E. K., Lauber, C. L., Hamady, M., Fierer, N., Gordon, J. I., and Knight, R. (2009). Bacterial community variation in human body habitats across space and time. *Science* 326, 1694–1697. doi: 10.1126/science.1177486
- Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. *Int. J. Complex Syst.* 1695.
- Ding, B., Gentleman, R., and Carey, V. (2014). *bioDist: Different Distance Measures*. R package version 1.36.0.
- Duran-Pinedo, A. E., Paster, B., Teles, R., and Frias-Lopez, J. (2011). Correlation network analysis applied to complex biofilm communities. *PLoS ONE* 6:e28438. doi: 10.1371/journal.pone.0028438
- Eiler, A., Heinrich, F., and Bertilsson, S. (2012). Coherent dynamics and association networks among lake bacterioplankton taxa. *ISME J.* 6, 330–342. doi: 10.1038/ismej.2011.113
- Faust, K., and Raes, J. (2012). Microbial interactions: from networks to models. *Nat. Rev. Microbiol.* 10, 538–550. doi: 10.1038/nrmicro2832
- Freilich, S., Kreimer, A., Meilijson, I., Gophna, U., Sharan, R., and Rupp, E. (2010). The large-scale organization of the bacterial network of ecological co-occurrence interactions. *Nucleic Acids Res.* 38, 3857–3868. doi: 10.1093/nar/gkq118
- Friedman, J., and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* 8:e1002687. doi: 10.1371/journal.pcbi.1002687
- Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature* 459, 193–199. doi: 10.1038/nature08058
- Fuhrman, J. A., and Steele, J. A. (2008). Community structure of marine bacterioplankton: patterns, networks, and relationships to function. *Aquat. Microb. Ecol.* 53, 69–81. doi: 10.3354/ame01222
- Gilbert, J. A., Steele, J. A., Caporaso, J. G., Steinbrück, L., Reeder, J., Temperton, B., et al. (2012). Defining seasonal marine microbial community dynamics. *ISME J.* 6, 298–308. doi: 10.1038/ismej.2011.107
- Girvan, M., and Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7821–7826. doi: 10.1073/pnas.122653799
- Goldfarb, K. C., Karaöz, U., Hanson, C. A., Santee, C. A., Bradford, M. A., Treseder, K. K., et al. (2011). Differential growth responses of soil bacterial taxa to carbon substrates of varying chemical recalcitrance. *Front. Microbiol.* 2:94. doi: 10.3389/fmicb.2011.00094
- HilleRisLambers, J., Adler, P. B., Harpole, W. S., Levine, J. M., and Mayfield, M. M. (2012). Rethinking community assembly through the lens of coexistence theory. *Annu. Rev. Ecol. Evol. Syst.* 43, 227–248. doi: 10.1146/annurev-ecolsys-110411-160411

- Horner-Devine, M. C., Silver, J. M., Leibold, M. A., Bohannan, B. J. M., Colwell, R. K., Fuhrman, J. A., et al. (2007). A comparison of taxon co-occurrence patterns for macro- and microorganisms. *Ecology* 88, 1345–1353. doi: 10.1890/06-0286
- King, A. J., Farrer, E. C., Suding, K. N., and Schmidt, S. K. (2012). Co-occurrence patterns of plants and soil bacteria in the high-alpine subnival zone track environmental harshness. *Front. Microbiol.* 3:347. doi: 10.3389/fmicb.2012.00347
- Kittelmann, S., Seedorf, H., Walters, W. A., Clemente, J. C., Knight, R., Gordon, J. I., et al. (2013). Simultaneous amplicon sequencing to explore co-occurrence patterns of bacterial, archaeal and eukaryotic microorganisms in rumen microbial communities. *PLoS ONE* 8:e47879. doi: 10.1371/journal.pone.0047879
- Koeppel, A. F., and Wu, M. (2012). Lineage-dependent ecological coherence in bacteria. *FEMS Microbiol. Ecol.* 81, 574–582. doi: 10.1111/j.1574-6941.2012.01387.x
- Langenheder, S., and Székely, A. J. (2011). Species sorting and neutral processes are both important during the initial assembly of bacterial communities. *ISME J.* 5, 1086–1094. doi: 10.1038/ismej.2010.207
- Larsen, P. E., Field, D., and Gilbert, J. A. (2012). Predicting bacterial community assemblages using an artificial neural network approach. *Nat. Methods* 9, 621–625. doi: 10.1038/nmeth.1975
- Laubert, C. L., Strickland, M. S., Bradford, M. A., and Fierer, N. (2008). The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biol. Biochem.* 40, 2407–2415. doi: 10.1016/j.soilbio.2008.05.021
- Leibold, M. A., and McPeck, M. A. (2006). Coexistence of the niche and neutral perspectives in community ecology. *Ecology* 87, 1399–1410. doi: 10.1890/0012-9658(2006)87[1399:COTNAN]2.0.CO;2
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., et al. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* 9:386. doi: 10.1186/1471-2105-9-386
- Newman, M. E. J., and Girvan, M. (2004). Finding and evaluating community structure in networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 69:026113. doi: 10.1103/PhysRevE.69.026113
- Oñiteru, I. D., Lunn, M., Curtis, T. P., Wells, G. F., Criddle, C. S., Francis, C. A., et al. (2010). Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15345–15350. doi: 10.1073/pnas.1000604107
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., et al. (2013). *vegan: Community Ecology Package*. R package version 2.0-10. Available online at: <http://CRAN.R-project.org/package=vegan>
- Philippot, L., Andersson, S. G. E., Battin, T. J., Prosser, J. I., Schimel, J. P., Whitman, W. B., et al. (2010). The ecological coherence of high bacterial taxonomic ranks. *Nat. Rev. Microbiol.* 8, 523–529. doi: 10.1038/nrmicro2367
- Pike, N. (2011). Using false discovery rates for multiple comparisons in ecology and evolution. *Method Ecol. Evol.* 2, 278–282. doi: 10.1111/j.2041-210X.2010.00061.x
- Proulx, S., Promislow, D., and Phillips, P. (2005). Network thinking in ecology and evolution. *Trends Ecol. Evol.* 20, 345–353. doi: 10.106/j.tree.2005.04.004
- Raes, J., and Bork, P. (2008). Molecular eco-systems biology: towards an understanding of community function. *Nat. Rev. Microbiol.* 6, 693–699. doi: 10.1038/nrmicro1935
- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., et al. (2011). Detecting novel associations in large data sets. *Science* 334, 1518–1524. doi: 10.1126/science.1205438
- Rivera, C. G., Vakil, R., and Bader, J. S. (2010). NeMo: network module identification in Cytoscape. *BMC Bioinformatics* 11:S61. doi: 10.1186/1471-2105-11-S1-S61
- Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat. Methods* 9, 811–814. doi: 10.1038/nmeth.2066
- Shade, A., Caporaso, J. G., Handelsman, J., Knight, R., and Fierer, N. (2013a). A meta-analysis of changes in bacterial and archaeal communities with time. *ISME J.* 7, 1493–1506. doi: 10.1038/ismej.2013.54
- Shade, A., Klimowicz, A. K., Spear, R. N., Linske, M., Donato, J. J., Hogan, C. S., et al. (2013b). Streptomycin application has no detectable effect on bacterial community structure in apple orchard soil. *Appl. Environ. Microbiol.* 79, 6617–6625. doi: 10.1128/AEM.02017-13
- Shange, R. S., Ankumah, R. O., Ibekwe, A. M., Zabawa, R., and Dowd, S. E. (2012). Distinct soil bacterial communities revealed under a diversely managed agroecosystem. *PLoS ONE* 7:e40338. doi: 10.1371/journal.pone.0040338
- Steele, J. A., Countway, P. D., Xia, L., Vigil, P. D., Beman, J. M., Kim, D. Y., et al. (2011). Marine bacterial, archaeal and protistan association networks reveal ecological linkages. *ISME J.* 5, 1414–1425. doi: 10.1038/ismej.2011.24
- Stegen, J. C., Lin, X., Konopka, A. E., and Fredrickson, J. K. (2012). Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J.* 6, 1653–1664. doi: 10.1038/ismej.2012.22
- Stone, L., and Roberts, A. (1990). The checkerboard score and species distributions. *Oecologia* 85, 74–79. doi: 10.1007/BF00317345
- Strimmer, K. (2008). fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. *Bioinformatics* 24, 1461–1462. doi: 10.1093/bioinformatics/btn209
- Sun, M. Y., Dafforn, K. A., Johnston, E. L., and Brown, M. V. (2013). Core sediment bacteria drive community response to anthropogenic contamination over multiple environmental gradients. *Environ. Microbiol.* 15, 2517–2531. doi: 10.1111/1462-2920.12133
- Tilman, D. (1982). *Resource Competition and Community Structure*. New Jersey, NJ: Princeton University Press.
- Walter, J., and Ley, R. (2011). The human gut microbiome: ecology and recent evolutionary changes. *Annu. Rev. Microbiol.* 65, 411–429. doi: 10.1146/annurev-micro-090110-102830
- Watts, D. J., and Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393, 440–442. doi: 10.1038/30918
- Zhalnina, K., de Quadros, P. D., Gano, K. A., Davis-Richardson, A., Fagen, J. R., Brown, C. T., et al. (2013). Ca. Nitrososphaera and Bradyrhizobium are inversely correlated and related to agricultural practices in long-term field experiments. *Front. Microbiol.* 4:104. doi: 10.3389/fmicb.2013.00104
- Zimmerman, A. E., Martiny, A. C., and Allison, S. D. (2013). Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *ISME J.* 7, 1187–1199. doi: 10.1038/ismej.2012.176

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 February 2014; accepted: 25 June 2014; published online: 18 July 2014.

Citation: Williams RJ, Howe A and Hofmockel KS (2014) Demonstrating microbial co-occurrence pattern analyses within and between ecosystems. *Front. Microbiol.* 5:358. doi: 10.3389/fmicb.2014.00358

This article was submitted to Terrestrial Microbiology, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Williams, Howe and Hofmockel. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.